

# Saliency-directed prioritization of visual data in wireless surveillance networks

Irfan Mehmood<sup>a</sup>, Muhammad Sajjad<sup>a</sup>, Waleed Ejaz<sup>b</sup>, Sung Wook Baik<sup>a,\*</sup>

<sup>a</sup> College of Electronics and Information Engineering, Sejong University, Seoul, Republic of Korea

<sup>b</sup> Department of Electrical and Computer Engineering, Queen's University, Kingston, Canada

irfanmehmood@sju.ac.kr, sajjad@sju.ac.kr, waleedeajz@sju.ac.kr, sbai@sejong.ac.kr

**Abstract:** In wireless visual sensor networks (WVSNs), streaming all imaging data is impractical due to resource constraints. Moreover, the sheer volume of surveillance videos causes intermittency on the analysts' ability to extract actionable intelligence from it. In this work, an energy-efficient image prioritization framework is presented to cope with the fragility of the traditional WVSN. The proposed framework selects semantically relevant information before transmission to the sink node. This is based on salient motion detection, which works on the principle of human cognitive processes. Each camera node estimates background bootstrapping, which helps to increase the efficiency of salient motion detection. Based on the salient motion, each sensor node is classified into a high or a low-priority node. This classification is dynamic such that camera nodes toggle between high-priority and low-priority status depending on the coverage of region of interest. The high-priority camera nodes are allowed to access the reliable radio channels for the sake of timely and reliable transmission of data. We compared the performance of this framework with other state-of-the-art methods in the cases of both single and multi-camera monitoring. The results indicate the usefulness of the proposed method in terms of salient event coverage and the reduction of computational and transmission costs as well as in helping analysts find semantically relevant visual information.

**Keywords:** image prioritization, wireless visual sensor networks, monitoring applications, salient activity detection.

## 1. Introduction

Wireless visual sensor networks (WVSNs) have recently emerged as a new type of sensor-based intelligent system. The performance and complexity challenges of the WVSNs transcend that of the existing wireless sensor

---

\* Corresponding Author

networks. In contrast to scalar sensors, WVSNs capture visual data, which offers rich information. Moreover, owing to their high versatility, small size, and dense spatial coverage, WVSNs can be deployed flexibly in various applications, such as in remote patient monitoring, distributed multimedia-based surveillance, and security systems [1-4]. In such setups, visual data about the event area has tremendous potential to influence decision making because the WVSNs address real-time observation of the events occurring within a busy environment. Despite the existence of numerous ideas of attractive surveillance applications, their actual implementation using the WVSNs remains a challenge. This is a challenge particularly because surveillance systems consist of numerous sensors that consume a large amount of bandwidth to transmit their raw video streams. Further, these video streams require extensive processing to detect events and anomalies. Recently, there have been attempts to advance the current WVSN technologies by implementation of intelligent methods for automatically extracting the relevant data at the source node to reduce the consumption of network bandwidth.

WVSN based monitoring systems have two main requirements: robustness and efficient resource utilization. In real surveillance environments, the robustness of the system is often compromised due to failure of camera sensors caused by unpredictable conditions such as natural catastrophic events, technical malfunctions, and human intrusion. A Multi-camera WVSN ensures the robustness of the system by providing visual access to the region of interest (ROI) even in the case of a few cameras failures. It has the capability to collaboratively share the monitoring tasks and minimize the energy consumption of each camera. However, a multi-camera wireless network introduces the problem of full or partial coverage overlaps in the camera's field-of-view (FoV) resulting in the production and transmission of highly redundant imaging data. Processing and transmission of such massive video data to the sink node causes unnecessary resource consumption of the whole network.

A wireless surveillance system is subject to a unique set of resource constraints such as finite on-board battery power and limited network communication bandwidth [1, 5]. Wireless transmission is close to its energy efficiency limit [6]. Energy consumption and the probability of packet collision increase in the case of transmitting all the multi-view data. Moreover, the huge volumes of visual data collected by the WVSN pose a critical hurdle for the human analysts by making it a challenge to discern actionable intelligence from it. Thus, it is imperative to leverage visual analytics at the point of data collection to analyze content from the video. This demands the development of intelligent and power-efficient methods for extracting semantically relevant data at the source node to reduce

network bandwidth consumption. These methods need to determine whether the sensed data is informative and should be transmitted or if it is better to wait for useful data before sending out any data. The decision must be made solely based on the coverage of the phenomenon being monitored.

In this paper, we have designed an energy-efficient visual data prioritization framework for multi-camera wireless networks based on image prioritization and its reliable transmission to the sink node. In the case of image prioritization, the camera sensors categorize visual content as either salient or non-salient. Saliency is a neurobiological notion that symbolizes the human ability to concentrate mental powers on certain areas by close observation and simultaneously disregard the irrelevant information [7]. For efficient monitoring, the proposed image prioritization calculates salient motion with a low-cost dynamic visual saliency model. It considers parameters such as motion contrast, the salient object's coverage ratio, and the user's preference to a specific camera. Moreover, the transmission of high-priority content to the sink node can be made error free by selecting a reliable channel for transmission. To ensure the selection of a reliable channel, cooperative spectrum sensing is introduced, in which the sink acts as a fusion center. This framework introduces a ranking strategy that formulates how cooperative sensing selects a reliable channel among the possible candidates in order to use the available bandwidth efficiently. This framework significantly prolongs the network lifetime and accelerates the decision making process by providing concise information at the base station (BS). In the rest of the paper, the terms "high-priority," "important," and "salient," while referring information are used interchangeably.

The main contributions of the proposed framework are as follows:

- A fully decentralized visual sensor-based intelligent system is proposed. It computes and disseminates aggregates of the multi-camera data with minimum processing and communication requirements.
- An energy-efficient image prioritization algorithm is presented using a low-cost dynamic visual saliency model. This extracts semantically important frames from the videos, thereby reducing the strain on the bandwidth.
- The dynamic visual saliency method is based on the fact that human visual attention is attracted by salient motion, which can be estimated using temporal gradients. To reduce the computation costs in modeling dynamic visual saliency, novel integral-image based temporal gradients are employed.

- Fully reliable data transmission is ensured by selecting the best available channel using cognitive radio networks.

The remainder of this paper is organized as follows. Section 2 and 3 summarize the related work and the proposed framework, respectively. Section 4 describes the evaluation scheme and experimental setup. The fifth and final section concludes the paper.

## 2. Related work

The development of automated video surveillance using the WVSNs has drawn considerable attention in recent years [8-12]. Aware of the limitations of the WVSNs and the dynamic nature of the surveillance environment, recent works has focused on certain aspects of visual monitoring, including collaborative data processing, visual data coverage, and visual data coding [13-15]. The main goal of the surveillance system is to use the least number of camera streams to avoid possible redundancy in the multimedia data while still providing the necessary event coverage. In this regard Dieber et al. [16] focused on the WVSN configuration problem. They presented an efficient approximation method based on an evolutionary algorithm to select an optimal configuration for cameras. However, in dealing with multi-camera video streams, these resource optimization schemes suffer because the computation and data distribution power have not been optimized for such an enormous amount of multi-view visual data.

To solve the problem of high power consumption while dealing with multi-camera wireless sensors, area coverage has been studied for efficient monitoring and optimal utilization of resources [17]. The authors in [18, 19] have proposed a solution for optimal FoV coverage. They selected the optimal subset of camera sensors for capturing an event area. Similar work is presented by Johnson and Bar-Noy on active cameras in [20]. They utilized the pan and scan camera properties to support full view coverage with a minimum number of cameras. Newell and Akkaya [21] proposed a heterogeneous sensor placement strategy that provides full coverage of the monitored areas using scalar and visual sensors. Other works in area coverage deal with the orientation of the camera sensors and occlusion issues for better coverage of the event area [22, 23]. However, these methods need to consider critical issues including efficient identification of the target location and coordination of the distributed sensors over the air, with limited resources. In addition, no mechanism is suggested to provide reliable data transmission, which is more crucial in WVSN based monitoring [24].

Mao et al. [25] mentioned that a majority of work in literature has focused on conventional data transmission methods without considering the resource-constrained dynamic nature of the WVSNs. For efficient transmission, Li et al. [26, 27] allocated more resources to the parts of the video sequence that have a greater impact on the video quality while spending fewer resources on those that are unimportant. In surveillance videos, humans are usually attracted towards the motion of the objects rather than the background [28]. Generally, the motion of objects is computed using video frame difference, background subtraction, and optical flow [29]. The frame difference scheme is simple and fast to implement, but it is not robust to various environmental challenges such as lighting changes, weather conditions (rain, storms) and dynamic background. Background subtraction methods perform efficiently in the case of static background but fail to adapt to the various background transformations. Similarly, optical flow detects motion in surveillance videos by considering the moving object speed, but it is not robust to illumination changes. To cope with these challenges various motion detection techniques have been presented in literature. Zahurul et al. [30] presented an improved and secure motion detection method for home and office environments. It incorporates both temporal change and optical flow detection methods. The major drawback in this scheme is that the selection of the optimal threshold hinders the implementation of the automatic surveillance system. A sift flow-based approach for modeling the background and detecting moving objects is presented in [31]. In this work, Dou and Li modeled each pixel as a group of adaptive Sift flow descriptors. These descriptors are computed over a rectangle region around the pixel, and the background model is dynamically updated. Huang and Cheng [32] presented a pyramidal background matching structure for motion detection in surveillance systems. This method utilizes spectral, spatial, and temporal features to generate a pyramid structure of the background model. Based on the background model, the motion of objects is detected by subtracting the estimated background model at each frame of the video sequence.

Unlike the aforementioned detection methods, researchers have recently focused on visual saliency-based salient motion detection approaches [33]. A real-time algorithm to detect salient motion in complex surveillance environments is presented in [34]. It combines temporal difference imaging and filtered optical flow to compute the salient motion. Su et al. [35] proposed a framework for solving the problem of abrupt motion tracking. This scheme integrates visual saliency model and particle filter tracker to cope with the motion tracking problems. Once the target is lost, a tracking recovery mechanism is implemented using a global search, which is realized by computing the saliency maps of the video frames. However, this method fails in environments with dynamic background

patterns. Lim et al. [36] introduced the idea of amplifying regions with salient motion in crowded scenes. It estimates the motion irregularities or anomalies by employing the dense optical flow algorithm and performs a two stage process to segment the salient regions. In addition, it eliminates the need to track each object individually and prior learning of the scene, which is important for real-time surveillance applications. The salient motion detection based mechanism is a better approach for the transmission of semantically important data. In this context, Irgan et al. [8] presented a salient region (important) based wireless video streaming system for packet prioritization that dynamically prioritizes image blocks using a cost-effective method. The salient micro-blocks are allowed to reach the sink node first for the sake of data completeness. It considers only single camera streams and does not provide any solution for reliable high-priority data transmission. Further, the problem becomes more difficult because of great data challenges in multi-camera wireless sensor networks [37]. Therefore, existing resource allocation methods fail to provide a solution for surveillance systems. In this work, we propose a salient motion detection-based energy-efficient visual data prioritization framework to cope with the fragility of the traditional multi-camera wireless sensor networks. A distributed approach is adopted in which all the nodes make autonomous measurements to prioritize the visual frames using the proposed salient motion detection scheme. These prioritization values are transmitted to the sink node. The sink node assigns appropriate transmission channels to each camera sensor based on periodic exchanges of sensor nodes' priority information and network statistics. Therefore, sensor nodes with lower priority have a non-reliable channel compared to the higher priority nodes. This will ensure the coverage of an event area significantly while reducing packet transmission costs. It will also help the analyst to find semantically relevant visual information.

### 3. Methodology

In multi-camera surveillance networks, it is necessary to consider and mitigate the impact of bandwidth limitations, distance, and network congestion to achieve reliable and efficient transmission of important video segments for human review. The proposed framework aims to design a cost-effective image prioritization method for the WVSNs that extracts semantically relevant content and efficiently transmits it to the BS. It is based on a distributed approach where camera nodes estimate visual saliency in order to prioritize video streams. The camera node status toggles between low and high-priority based on salient motion detection. The sink node assigns a high-priority sensor node a reliable channel as compared to the low-priority nodes. This ensures the coverage of an event

area, while significantly reducing the network packet transmission cost. Figure 1 illustrates the main steps of the proposed framework.

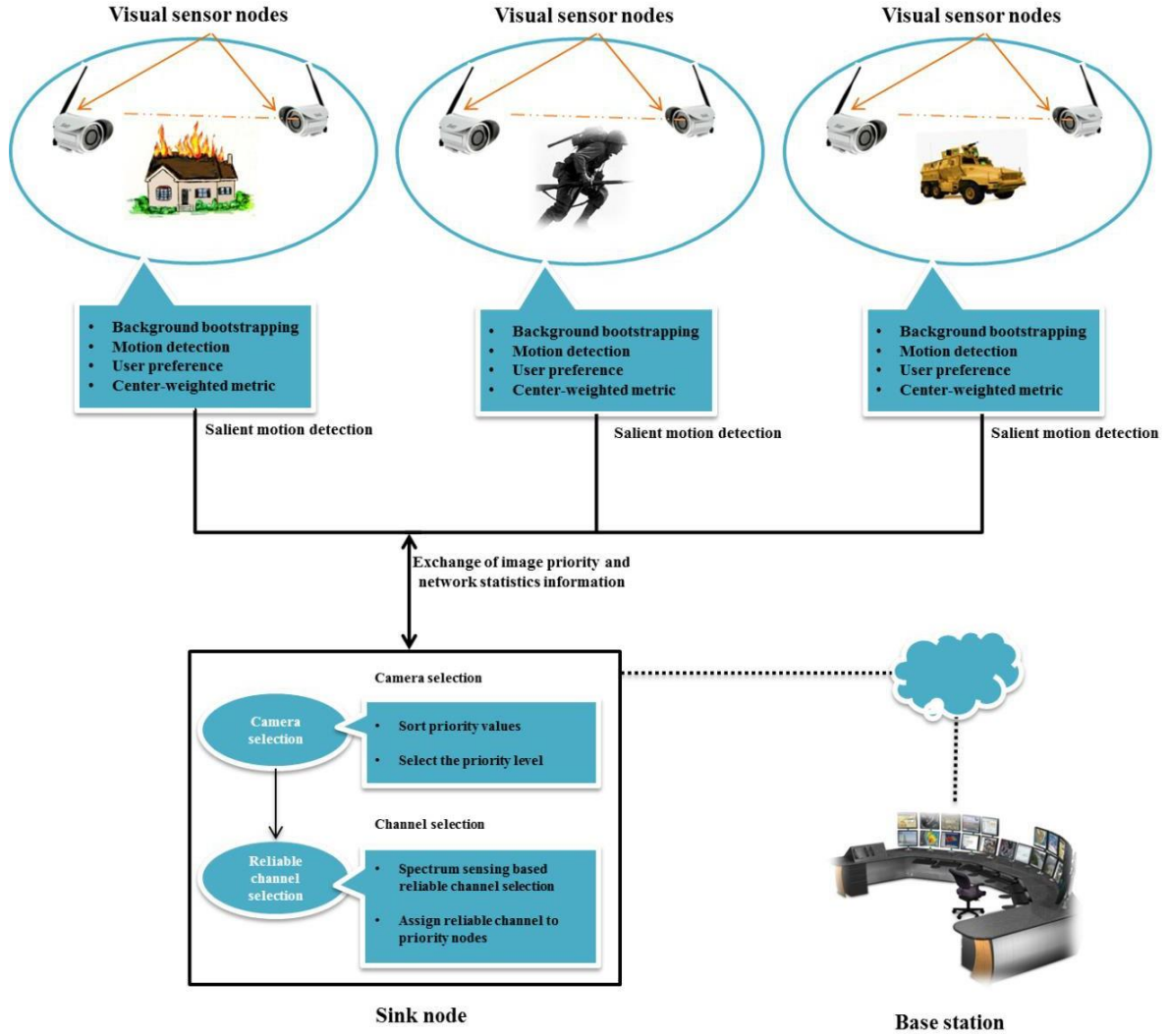


Fig. 1: Proposed framework for image prioritization in multi-camera wireless visual surveillance networks.

Consider a surveillance system consisting of numerous wireless sensor cameras located in an area of interest (e.g. border) to monitor activities and to transmit required information to the sink node using a wireless network. Each sensor node has an image processor and a transceiver module. Image data from the sensor is processed by the processing module. The on-board image / vision processor has the abilities of low-level image processing operations, e.g. edge detection, motion detection using kernel operations like convolutions. A transceiver module is used for communicating with other cameras and the sink node.

### **A. Network agents**

WVSNS' agents are classified as follows:

#### **1. High-priority camera nodes ( $N^{PH}$ ):**

The high-priority cameras detect regions of coherent motion (salient motion) and transmit video via reliable channels. The regions of coherent motion are computed with energy-efficient integral-image based temporal gradients. This creates an intelligent camera node that is capable of determining whether the sensed visual data is important or not.

#### **2. Low-priority camera nodes ( $N^{PL}$ ):**

The extraction of low-level features in a camera node plays an important role in determining the priority of a specific node. A low-priority is assigned to a camera node when the camera's FoV does not detect any salient activity. Therefore, a node having a low priority waits and observes the environment to detect some salient event before sending out any data. Camera nodes having a low priority act as relay nodes and only forward high-priority video data through the wireless mesh network to the sink node.

#### **3. Sink node (SN):**

The SN queries all other nodes in its domain and acts as a data fusion center. Whenever an event is generated, the camera sensors broadcast the visual information to the SN for it to take appropriate action. Based on the received image priority, the SN adaptively assigns transmission channels according to the required reliability.

### **B. Energy-efficient image prioritization**

Salient motion detection is very important in video surveillance because it focuses on interesting events; this increases the effectiveness of surveillance and reduces the cognitive overload on the human operator. However, in complex surveillance environments (for example at the borders, in forests, and in public places) motion can be coherent or non-coherent. Salient motion is defined as a coherent motion that is likely to result from typical surveillance objects such as a person or a vehicle as opposed to other motions, like that of tree branches, grass, and the oscillation of vegetation in the wind. For such environments, we present a real-time algorithm designed to detect the salient motion and suppress the non-salient motion. It is a low-cost algorithm that can be implemented in a wide range of small, low-power devices. The key principle that provides salient motion detection in real time is the use of



novel features derived from an integral-image. It was introduced in the field of Computer Graphics for texture mapping [38]. The integral-image has a value at a pixel  $(x, y)$  equal to the sum of pixels above and to the left of it, inclusive. Later, Viola and Jones [39] used this idea for object detection, and they proved that integral-image produces results 15 times as fast as previous works.

Initially, integral-image is calculated from each video frame captured by the visual sensor nodes. Then, background bootstrapping is done using the integral-image. Consider two consecutive video frames  $F_n^t$  and  $F_n^{t-\Delta t}$  captured by the  $n^{\text{th}}$  camera sensor at time  $t$  and  $t-\Delta t$  respectively; their corresponding integral-images at location  $(x, y)$  are calculated as follows:

$$\mathcal{F}_n^t(x, y) = \sum_{x' \leq x} \sum_{y' \leq y} F_n^t(x', y'); \quad x \in W, y \in H \quad (1)$$

$$\mathcal{F}_n^{t-\Delta t}(x, y) = \sum_{x' \leq x} \sum_{y' \leq y} F_n^{t-\Delta t}(x', y'); \quad x \in W, y \in H \quad (2)$$

where  $W$  and  $H$  are the width and height of the image. It can be conceded from equation 1 and 2 that integral image is the sum of pixels within the rectangular area from the upper left point to the lower right point. An integral-image (or summed area table) is a lightweight process that can be estimated in one pass over the original image. An example of an integral-image is shown in figure 2. A video frame  $F_n^t$  in Figure 2a shows the activity of soldiers in combat filed and its respective image divided into blocks of size  $N \times N$  pixels as given in figure 2b. Figure 2c shows the integral-image map  $\mathcal{F}_n^t$  with each pixel showing the value of the integral-image of that point. Consider a rectangular area  $D$  of image  $F_n^t$  defined by  $P_1 = (x1, y1)$  and  $P_2 = (x2, y2)$  as shown in figure 2c, its sum can be calculated using *rectSum* function as

$$\text{rectSum}(D) = \text{Area}(A+B+C+D) - \text{Area}(A+B) - \text{Area}(A+C) + \text{Area}(A) \quad (3)$$

In terms of integral-image  $\mathcal{F}_n^t$ , equation (3) can be written as:

$$\text{rectSum}(D) = \mathcal{F}_n^t(x2, y2) - \mathcal{F}_n^t(x1, y2) - \mathcal{F}_n^t(x2, y1) + \mathcal{F}_n^t(x1, y1) \quad (4)$$

The above equation can also be expressed as

$$rectSum(F_n^t(x1, y1, x2, y2)) = \mathcal{F}_n^t(x2, y2) - \mathcal{F}_n^t(x1, y2) - \mathcal{F}_n^t(x2, y1) + \mathcal{F}_n^t(x1, y1) \quad (5)$$

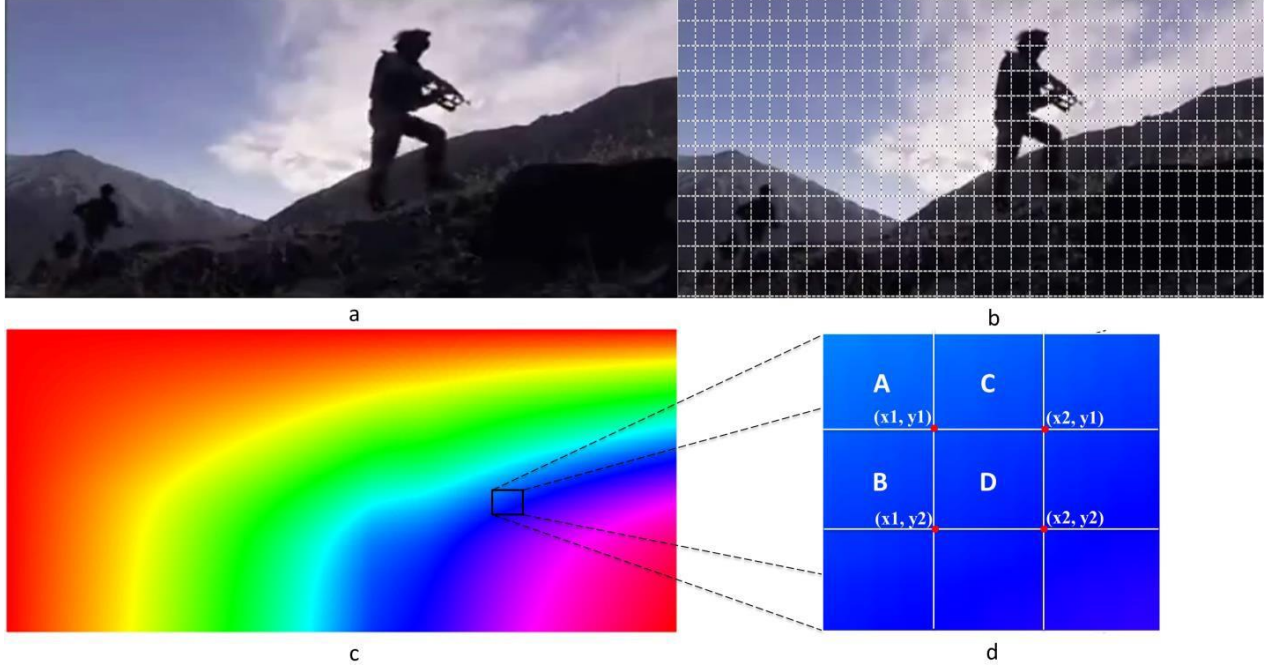


Fig. 2: Integral-image and evaluation over a rectangle. (a) Border surveillance test image, (b) Macro-blocks of the image, (c) Integral-image map, (d) Computing the sum of pixel values in rectangular region.

The background model (bootstrapping)  $\mathcal{B}_n^m$  for  $n^{\text{th}}$  camera is updated whenever there is a sufficiently high correlation between two consecutive integral frames. Thus, for background bootstrapping, first we compute the correlation coefficient  $CC^B$  between two macro-blocks  $A_{t-\Delta t}$  and  $A_t$  as

$$CC^B(i, j) = \frac{\left| rectSum(F_n^{t-\Delta t}(x1, y1, x2, y2)) - \mu_{A_{t-\Delta t}} \right| * \left| rectSum(F_n^t(x1, y1, x2, y2)) - \mu_{A_t} \right|}{\sqrt{\left| rectSum(F_n^{t-\Delta t}(x1, y1, x2, y2)) - \mu_{A_{t-\Delta t}} \right|^2 * \left| rectSum(F_n^t(x1, y1, x2, y2)) - \mu_{A_t} \right|^2}} \quad (6)$$

Here  $\mu_{A_{t-\Delta t}}$  and  $\mu_{A_t}$  are the average pixel values of macro-blocks  $A_{t-\Delta t}$  and  $A_t$ , respectively. The initial background bootstrapping is ready when all the blocks have been set to stable state (i.e. when there is no significant background motion). These steps are essential to remove the background motion and to accurately estimate the salient motion.

In order to obtain the motion information in video sequences quickly, the concept of integral-image based temporal gradients is presented to help compute salient motion implicitly by measuring the changes in the values of image blocks in neighboring frames, which approximate the motion component of visual attention. The computation of integral-image based temporal gradients is considerably cheaper than traditional methods like simple temporal gradients, and optical flow [40, 41]. Motion saliency (SM) is computed by calculating the derivatives at each block. For static points, we use the updated background model  $\mathcal{B}_n^m$ , or otherwise we use video frame  $\mathcal{F}_n^t$ . Salient motion

$SM_n^t$  captured by the  $n^{\text{th}}$  camera sensor at time  $t$  is computed as follows:

$$\left. \begin{aligned} SM_n^{t,x}(x, y) &= \frac{\partial F_n^t(x, y)}{\partial(x, t)} \\ &= rectSum(F_n^t(x1-1, y1, x2-1, y2)) - rectSum(F_n^t(x1+1, y1, x2+1, y2)) \\ SM_n^{t,y}(x, y) &= \frac{\partial F_n^t(x, y)}{\partial(y, t)} \\ &= rectSum(F_n^t(x1, y1-1, x2, y2-1)) - rectSum(F_n^t(x1, y1+1, x2, y2+1)) \end{aligned} \right\} \quad (7)$$

$$\left. \begin{aligned} SM_n^{t,x}(x, y) &= \frac{\partial F_n^t(x, y)}{\partial(x, t)} \\ &= rectSum(F_n^t(x1-1, y1, x2-1, y2)) - rectSum(\mathcal{B}_n^m(x1+1, y1, x2+1, y2)) \\ SM_n^{t,y}(x, y) &= \frac{\partial F_n^t(x, y)}{\partial(y, t)} \\ &= rectSum(F_n^t(x1, y1-1, x2, y2-1)) - rectSum(\mathcal{B}_n^m(x1, y1+1, x2, y2+1)) \end{aligned} \right\} \quad (8)$$

Salient motion  $SM_n^t$  is then obtained as

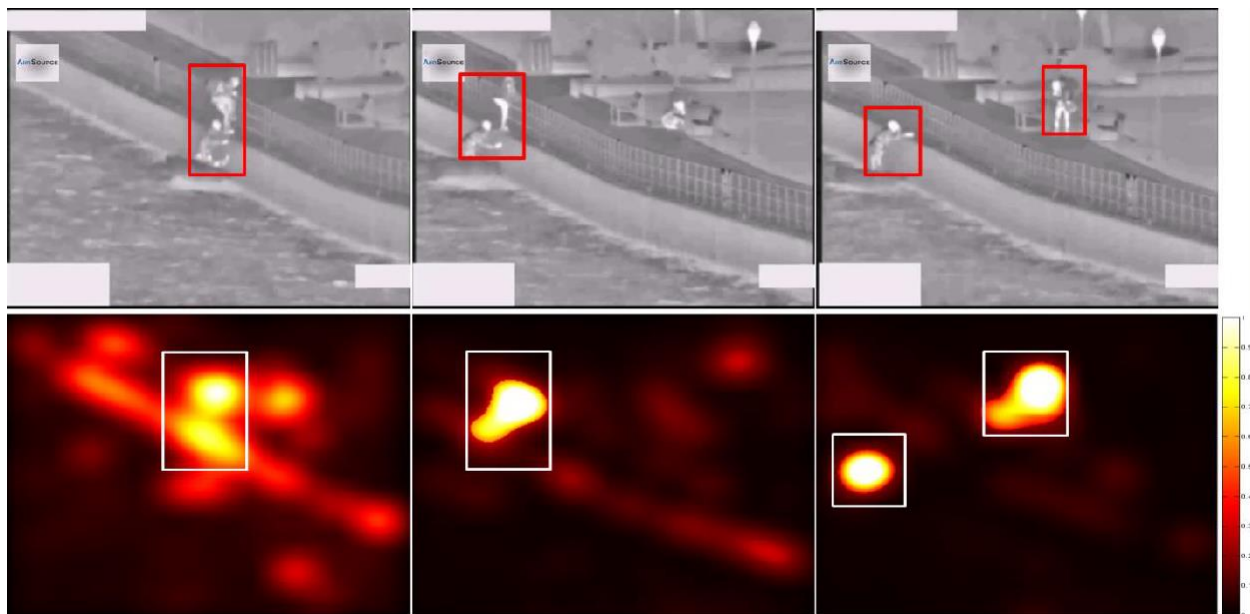
$$SM_n^t = \sqrt{\frac{\partial F_n^t(x, y)^2}{\partial(x, t)} + \frac{\partial F_n^t(x, y)^2}{\partial(y, t)}} \quad (9)$$

Salient motion  $SM_n^t$  computed by equations 7 and 8 is quite robust to small motions in the background, mainly

because of the use of the background model  $\mathcal{B}_n^m$  and the integral-image based temporal gradients, as shown in

figure 3. We present a descriptive analysis of a potential example related to the extraction of semantically related information using the above mentioned salient motion detection algorithm from a surveillance video. The underlying

thermal video-shot consists of 1427 frames that capture illegal border crossing activity along the *St. Clair River in Detroit*. The video-shot was captured by *Kris Grogan / U.S. Customs and Border Protection Office of Public Affairs - Visual Communications Division* using a remote video surveillance system camera. The video was downloaded from AiirSource<sup>1</sup>, which specializes in current and archival military/aviation videos from U.S. Army, Navy, Marine Corps, Air Force, and Coast Guard. In this evaluation, regions with salient motion should be correctly detected referring to suspicious activity. The salient motion is given by a group of people climbing and crossing the border fence at a river bank as shown in Figure 3. To highlight the salient motion, the people crossing the fence in frames are enclosed by the red rectangle. The frame sequence in this video shows a significant amount of motion clutter due to the river waves and the strong wind shaking the camera, which in turn is continuously changing the patterns of the background. These distracting motions make salient motion detection more challenging. However, from Figure 3 it can be easily conceded that the proposed salient motion detection technique performs well even in the most challenging environments. Figure 3 shows the results for the salient motion detection in the underlying video by the proposed method. The top row shows sample frames taken from the video, and the bottom row highlights the salient motion detection. It can be easily concluded that the salient motion due to the illegal crossing of the border fence has been reliably detected. This salient motion detection can help users quickly locate a semantically relevant position in frames sequence captured by visual sensors.



<sup>1</sup> <https://www.youtube.com/watch?v=G68Lr2LAjYs>

Fig. 3: While monitoring a border fence along the bank of a river, security officials are interested in detecting people crossing/climbing the fence. Top row: sequence of thermal video frames showing a group of people crossing the fence. Bottom row: motion saliency map computed by the proposed technique highlights the object owing salient motion.

There are certain general principles whose implementation results in the improved understanding of the scene with a more efficient monitoring. Two such important factors, center-weighted metric  $\alpha$  and user preference  $\mathcal{U}$  are incorporated to estimate the semantic and narrative requirements of a salient event in the scene. Human beings tend to focus more on the center of the scene [42]. Therefore, a simple center-weighted mechanism is used to assign high weights to the salient objects in the middle of the frame. For this purpose, the initial Euclidean distance of each pixel is calculated from the central pixel. Then, a Gaussian function is used to assign weights in order to obtain a center-weighted metric  $\alpha$  [43] as given

$$\alpha = e^{-\left(\frac{\sqrt{(x-Cx)^2+(y-Cy)^2}}{\sigma}\right)^r} \quad (10)$$

where  $\sigma = c * \sqrt{(W/2)^2 + (H/2)^2}$  and  $(Cx, Cy)$  is the center position of the image. The prioritization value  $\mathcal{P}_n^t$  for each video stream is obtained using  $\alpha$ , and the user preference  $\mathcal{U}$  for that specific camera is given.

$$\mathcal{P}_n^t = (\alpha * SM_n^t) * \mathcal{U}^n \quad (11)$$

where  $\mathcal{U}^n \in [0, 1]$  is the pre-assigned interest of a user in a particular surveillance area covered by camera node  $n$ .

Based on the priority value  $\mathcal{P}_n^t$ , each sensor node is classified into high or low-priority nodes. The classification is dynamic where camera nodes toggle between high-priority and low-priority status, depending on the information captured.

### C. Cooperative sensing based channel selection

In order to realize efficient multimedia communication in WVSAN, it is imperative to have a reliable transport mechanism. Using a dedicated spectrum for multimedia communication in WVSAN is difficult because of congested bandwidth allocation. Cognitive radio can help to improve overall spectrum utilization and to preserve the limited

resources of wireless sensor networks [44]. Cognitive radio enabled sensor networks can access a spectrum opportunistically, and a new paradigm of sensor networks is formed, which is called the cognitive radio multimedia sensor networks (CRMSN). Therefore, a delay in sensitive multimedia applications in CRMSN can be realized by the addition of low cost hardware for cognitive capabilities [45]. Recent research shows that spectrum issues in WVSNS can be alleviated by incorporating cognitive radio (CR) technology [46, 47]. CR is a key technology that can help to mitigate the scarcity of the spectrum by using licensed spectrum bands opportunistically for unlicensed users. The advantages of CR that are associated with the opportunistic access of the unused licensed band include the alleviation of spectrum scarcity, large bandwidth, long-range communication using TV band, and reduced cost for communication.

Spectrum sensing is required to detect a licensed user's activity. A high probability of detection and low probability of false alarm is required in order to satisfy all the requirements of cognitive radio networks. The performance of spectrum sensing can be improved by cooperation among the cognitive users. To fulfill this purpose, a cooperative spectrum sensing scheme is proposed for the multi-camera surveillance system. Linear m-out-of-n fusion rules are applied to arrive at a cooperative decision, in which n is the number of users participating in the cooperation. The sink will act as a fusion center in the multi-camera surveillance system and will forward the cooperative decision to all the individual cameras [48]. However, the reporting time increases with increase in cooperative cameras in the spectrum sensing process, which results in a lower network throughput. Therefore, the optimal number of cooperative cameras and the m-out-of-n rule are required to satisfy the detection performance constraints.

Let the local decision of the  $i^{\text{th}}$  camera be  $LD_i$ , which can be one or zero with respect to the presence or absence of the licensed user, respectively. The cooperative decision at the sink is made by performing the logical AND on all the local decisions reported by cognitive radio enabled cameras in the AND rule. The cooperative decision about the presence of a licensed user is made only when all the cognitive radio enabled cameras report the presence of the licensed user, while the cooperative decision about the absence of the licensed user is made when any one of the cognitive radio enabled cameras reports an absence. The logical OR is performed in the OR rule at the sink on all the local decisions made by cognitive radio enabled cameras. The cooperative decision about the absence of the licensed user is made only when all the cameras report the absence of the licensed user, and the cooperative decision about licensed user's presence is made when any one of the cognitive radios enabled cameras reports the licensed

user's presence. The majority rule requires at least half of the cognitive radio enabled cameras to report the presence of licensed users to confirm the presence and vice versa.

The generalized form of these simple rules uses m-out-of-n rules in which the sink declares a cooperative decision about licensed users' presence, if more than m cognitive radio enabled cameras among n cognitive radio enabled cameras report the licensed user presence. The global probability of false alarm and detection is calculated at the central entity by fusing the results of cognitive radio enabled cameras. The global probability of detection  $Q_d$ , and the global probability of false alarms  $Q_f$ , for the m-out-of-n rules are given by Zhang et al.[49]:

$$Q_f = \sum_{j=m}^n \binom{n}{j} P_f^j (1 - P_f)^{n-j} \quad (12)$$

$$Q_d = \sum_{j=m}^n \binom{n}{j} P_d^j (1 - P_d)^{n-j} \quad (13)$$

where  $P_d$  and  $P_f$  are the probability of detection and the probability of a false alarm determined by each cognitive radio enabled camera. The total number of cognitive radio enabled cameras participating in cooperation is n, and m determines the rule among AND, OR and majority.

The flow of the proposed cooperative spectrum sensing algorithm for the multi-camera surveillance system is shown in Figure 4. In this framework, the relay node is used instead of the traditional first-in-first-out packet transmission, giving privilege to high-priority nodes, allowing them to be forwarded before other packets. Thus, the priority packets are transmitted reliably to the SN in a timely manner. There are n cognitive radio enabled cameras participating in the proposed cooperative spectrum sensing for multi-camera surveillance system. Each camera performs energy detection for local spectrum sensing [49] and sends the decision to the sink. The sink acts as a fusion center, is capable of measuring the channel conditions, and ranks the available channels from the results of spectrum sensing according to channel quality parameters. Finally, it assigns the most reliable channels to high-priority sensors for data transmission.

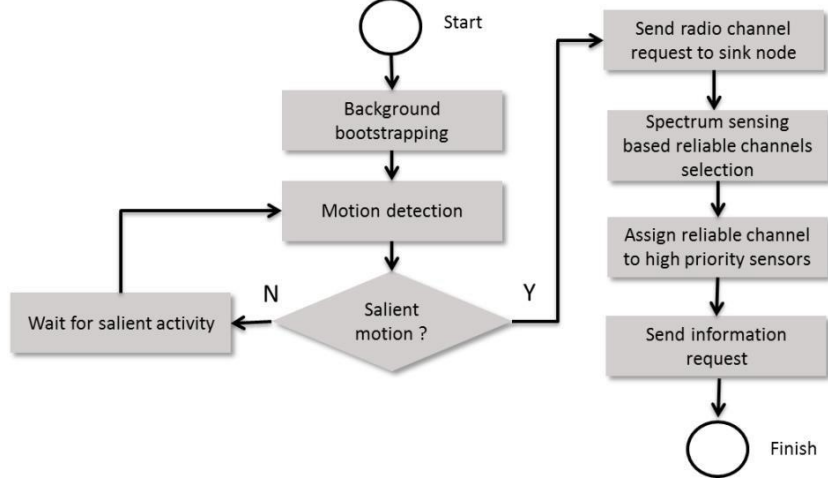


Fig. 4: Flowchart of the high-priority camera selection and reliable packet transmission algorithm for multi-camera surveillance networks.

#### 4. Evaluation and experimental setup

In order to simulate the proposed method, Tmote Sky and Matlab are used. We compared the results of the proposed framework with state-of-the-art prioritization schemes [8, 34-36]. For a comparative analysis, the Tmote Sky used in this simulation has similar specifications as the Irgan method [8]. The selection of a video stream out of multiple videos is based on a priority measure that computes the level of salient activity in the environment. The priorities of multi-camera video sequences are calculated using the Matlab software package. The priority value of whole image is mapped to its macro-blocks (size  $8 \times 8$ ). These macro-blocks are transmitted using channels ranked by cognitive radio networks that support the transmission of a macro-block (66 bytes including two bytes macro-block offset) in a data packet. All the packets that belong to the high-priority image are transmitted via a reliable channel, and the low-priority packets are dropped.

We conducted two sets of experiments on surveillance videos acquired from indoor and outdoor multi-cameras. These experiments were aimed at assessing the capability of the proposed algorithm for correctly detecting the salient activities in the monitored scene and classifying these activities as either high-priority or low-priority. For this purpose, standard data sets were used, as mentioned in Table 1. The main criterion for data sets selection was their suitability to surveillance applications. Indoor multi-cameras monitor incidents in a moving train. In these video sequences, several incidents were generated. Simulations were performed on the footages of four cameras consisting of activities such as mobile theft, fire, and passengers' fighting. Outdoor multi-cameras observe the



activities of an airplane and vehicles moving across some roads around a runway. In addition, it observes action such as entering a building and parking a car.

No.	Sequence name	Sequence description	# of frames	Category
1	Sequence 1	Theft of a mobile phone.	1400~1600	<sup>2</sup> Indoor Multi-camera videos: The sequences record incident situations in the real conditions of a moving train. All sequences are shot by ten cameras
2	Sequence 2	Passenger staring at another man's wife, then a fight ensues.	3500~3950	
3	Sequence 10	Passenger notices a fire in the next coach, and everybody runs out of the train.	2500~2775	
4	Sequence 6	Man harasses a woman.	2600~2950	
5	ETI-VS2-AP	Shows an airplane and other vehicles moving across some roads around runway.	700~804	<sup>3</sup> Outdoor Multi-camera Videos: These sequences record the activities of persons and other objects like vehicles in different regions of interest. All sequences are shot by four cameras
6	ETI-VS2-BE-19	Observes objects performing different actions for example entrance in the building, and parking a car, etc.	720~1025	

Table 1: Details of test videos, downloaded from live databases<sup>1,2</sup>.

The evaluation of image prioritization methods is a contentious issue because there is no standard evaluation methodology. This is mainly due to a lack of an objective ground truth and the subjective nature of the definition “high and low-priority image.” Therefore, three evaluation criteria were used.

#### 4.1. Significance of the proposed method

The first evaluation mechanism is based on a test case to provide an insight into the proposed technique. The results have been presented on single shots that were captured in the ETI-VS2-BE-19 video sequence by two cameras. In these shots, a person is walking out of a building and this activity is monitored by two cameras located inside and outside the building. The priority curves of the salient motions of each camera are shown in Figure 5. Visual frames at the top of the priority curves possess high-priority contents. The frames at the bottom of the curves indicate low-priority video segments, which do not possess any salient activity. In the initial part of the shot, it is evident from the priority curves that visual activity captured by the ETI-VS2-BE-19-C4 camera has a greater high-priority value than ETI-VS2-BE-19-C2. This is because the person is going out of the building, and after he left the building there was

<sup>2</sup><https://www.sop.inria.fr/orion/ETISEO/videos/>

<sup>3</sup><http://www.multitel.be/image/research-development/research-projects/boss.php>

no salient activity in ETI-VS2-BE-19-C4. On the other hand, the visual activity occurring in the ETI-VS2-BE-19-C2 camera coverage area is high during the middle of the shot, where it captures a person coming out of the building and walking away. It can be seen from the priority curve of ETI-VS2-BE-19-C2 that its value decreases as the person walks out of its coverage area. Here, the video streams having important visual content are selected and transmitted to the SN. In this way, users get the most succinct information in a timely manner. This priority scheme selects the semantically important activities and thus reduces the burden of video analysis by a significant ratio.

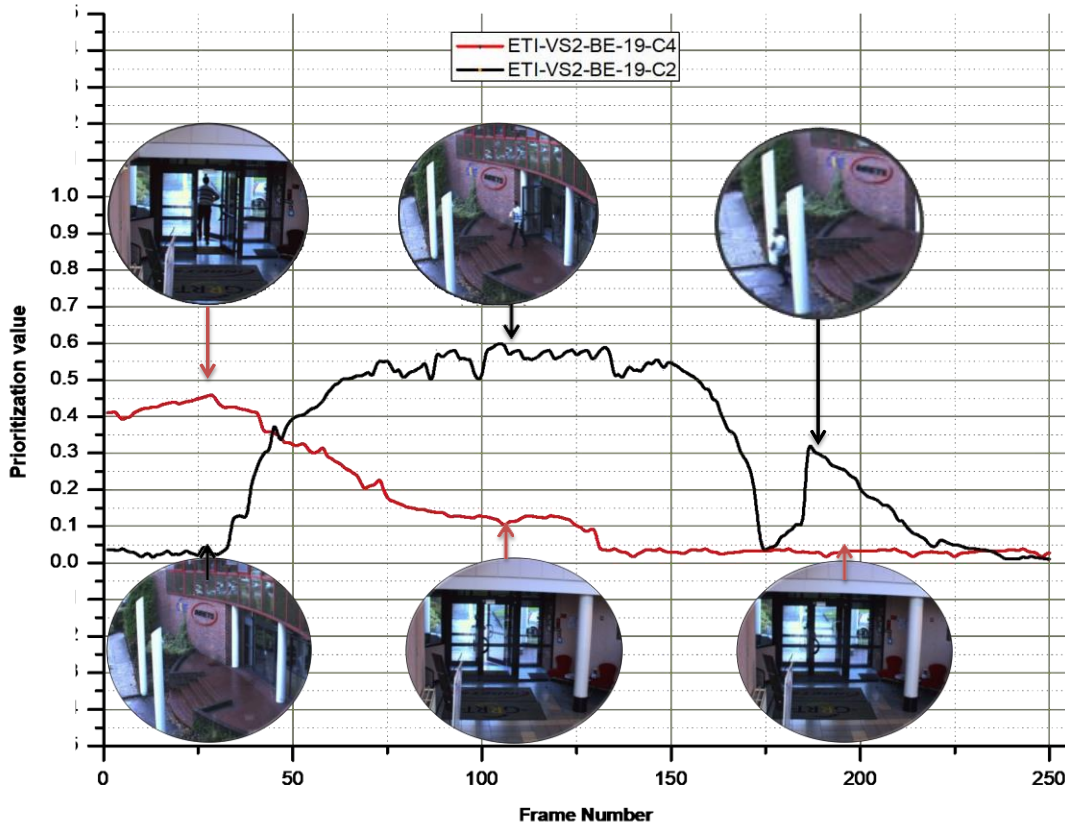


Fig. 5: Priority curves for the video sequence ETI-VS2-BE-19 showing the selection of high-priority frames in a simple shot.

#### 4.2. Robustness to various challenges in real-world environment

In the neurosciences, it has been observed that salient motion detection is a compelling way of focusing on relevant information and suppressing irrelevant data [50]. Salient motion detection is an application-dependent process that separates relevant information from non-relevant information. In a real-world environment, motion can be caused by

both salient (interesting) objects and non-salient (uninteresting or distracting) objects. The non-salient objects such as swaying leaves, grass etc. produce distracting motions that make the problem of salient motion detection more challenging. In this section, the robustness of the proposed salient motion detection algorithm was tested on four videos related to different categories as described in Table 2. These videos were downloaded from the open database, *Change detection* [51], which is a standard dataset for the evaluation of motion detection techniques. The selected videos cover a wide range of detection challenges and are representative of typical visual data captured in surveillance scenarios. Each video frame is accompanied by accurate ground-truth segmentation and annotation of change/motion areas.

No.	Category	Video name	Video description	# of frames
1	Dynamic background	Fountain02	Moving car with background motion of fountain and swaying leaves.	1499
2	Bad weather	Blizzard	Road traffic in bad weather (blizzard).	7000
3	Pan, Tilt, and Zoom (PTZ)	ZoomInZoomOut	Activities of man with PTZ camera.	1130
4	Turbulence	Turbulence3	A motorcyclist and vehicles are crossing a river bridge with extreme turbulence flow caused by the wind and river.	2200

Table 2: Details of test videos, downloaded from *Change detection* [51] database.

In Figure 6, the top row shows three sample frames of the video, *fountain02.mpeg*, where a car (enclosed by a red rectangle) moves from right to left. The motion of car is salient with the intuition that human visual system focuses on visual persistent motion. This video exhibits an animated background due to the background moving objects (fountains and swaying leaves). The background model (bootstrapping step) of the proposed salient motion detection algorithm efficiently adapts to background changes. Both background model and integral-image based temporal gradients efficiently extract the coherent motion (that is, motion of car) and discards the inconsistent or flickering motion (that is, motion of water fountains and leaves). The second video, *Blizzard.mpeg*, contains road traffic monitoring of various rapid moving vehicles on a tow-way road accompanied by snowfall and strong winds as shown in Figure 7. The main challenges are snow storm and blurring of FOV that reduce clarity of images causing false alarms in motion detection. The proposed scheme solves this problem effectively by continuously updating the background model and computing the integral-image based temporal gradients. These temporal gradients help to highlight the motion of salient objects in the video (moving vehicles) as shown in Figure 7.

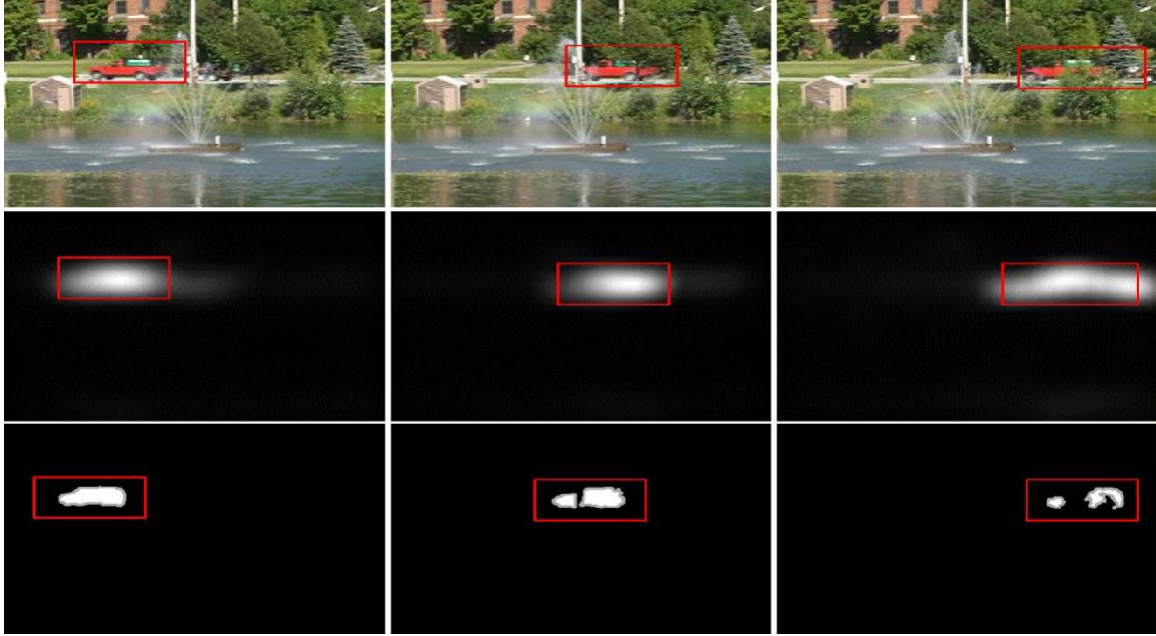


Fig. 6: Performance evaluation of the proposed salient motion detection algorithm in a dynamic background. First row shows original image sequence, red rectangle shows the object with salient motion. Second row shows salient motion detection by the proposed method. Third row shows the ground truth for salient object motion detection.



Fig. 7: Performance evaluation of the proposed salient motion detection algorithm in a video with widespread snowfall accompanied by strong winds and vehicles are crossing the roads. First row shows original image

sequence, red rectangle shows the object with salient motion. Second row shows salient motion detection by the proposed method. Third row shows the ground truth for salient object motion detection.

Nowadays, PTZ cameras are widely used for surveillance purposes. Most of these PTZ cameras are preprogrammed such that they can be moved along pre-defined trajectories. However, it would be more effective if the PTZ camera could automatically detect, follow, and record salient activities. To check the possibility of the proposed method allowing PTZ cameras to automatically localize these salient activities, we tested it on *ZoomInZoomOut.mpeg* video. This video was captured with a PTZ camera in which camera is gradually zooming out as the person walks as shown in Figure 8. As the PTZ camera *zooms in* or *zooms out*, the background model updates itself and integral-image based gradients utilize this updated background model for motion detection of the walking person. The middle row in Figure 8 shows the salient motion detection results, which seems to be effective in focusing the salient motion in the scene. Figure 9 shows visual frames of the *Turbulence3.mpeg* video. In this video shot a motorcyclist is passing over the bridge with strong turbulence because of wind and river waves. These turbulent flows cause highly irregular and distracting motion patterns and impose challenges in detecting and tracking the salient motion (the motorcyclist). Usually, the motion of turbulent flows is characterized by distracting and incoherent structures because these motions are inconsistently diverging in different directions. Our motion detection method relies on efficient background modeling and temporal gradient estimation between consecutive frames and is more suitable to be used for estimating the movement of objects in turbulence environments. In this case, integral-image based temporal gradients effectively separate the inconsistent motion patterns and highlight the salient motion patterns. From Figure 9, it can be easily conceded that the proposed algorithm efficiently detects the salient motion of the motorcyclist in highly challenging turbulence conditions. Thus, despite the complex dynamic content and various challenges offered by the real surveillance environments, our algorithm reliably detects the salient motion. Based on the salient motion discussed earlier, each sensor node is classified into a high or a low-priority node.

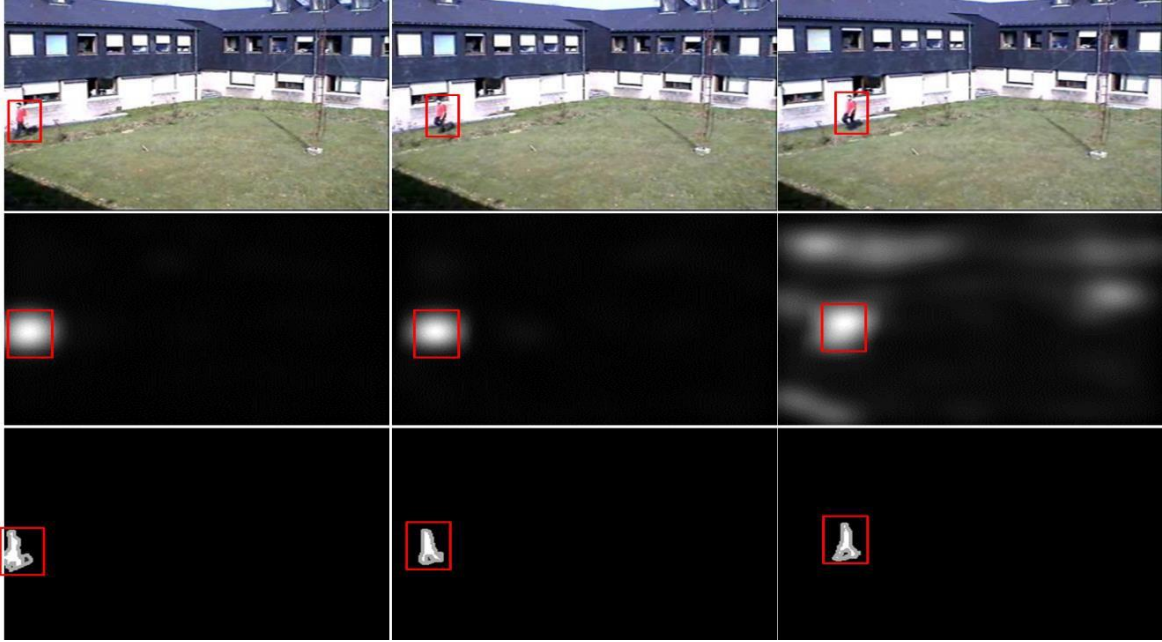


Fig. 8: Performance evaluation of the proposed salient motion detection in the video *ZoomInZoomOut.mpeg*, showing a person walking in front of a building while camera is gradually zooming out. First row shows original image sequence, red rectangle shows the person with salient motion for display purpose. In the second row, the red rectangle shows that the proposed salient motion detection accurately detects objects with salient motion. Third row shows the ground truth for salient object motion detection.

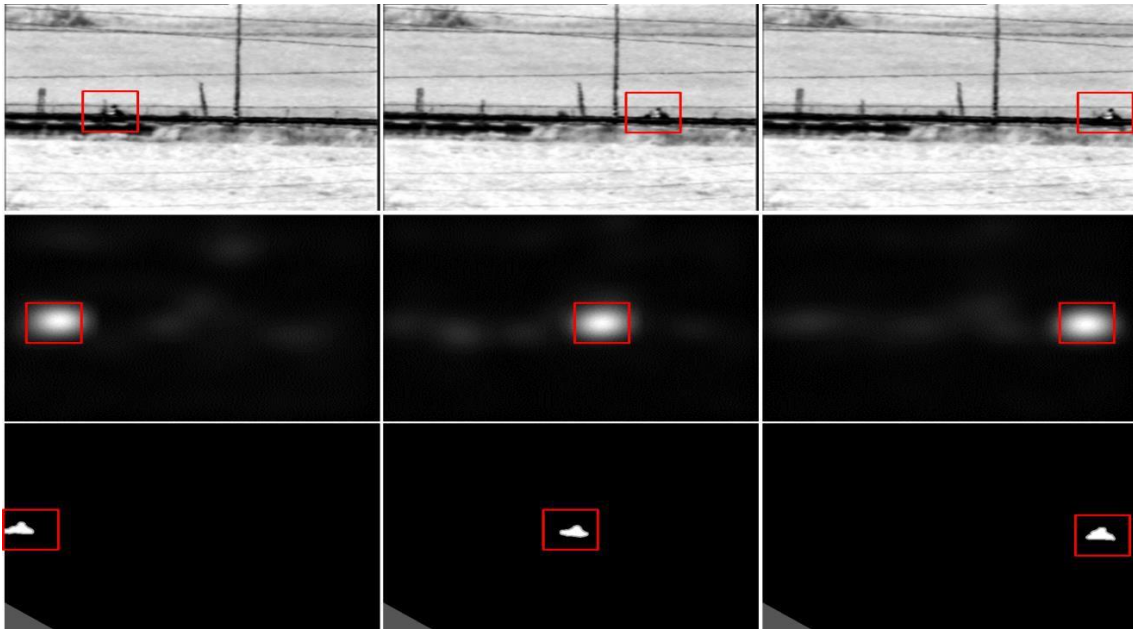


Fig. 9: Performance evaluation of the proposed salient motion detection in the video *Turbulence3.mpeg*, showing a motorcycle is passing by the bridge with strong air and river wave's turbulence. First row shows original image sequence, red rectangle shows the object with salient motion. Second row shows salient motion detection by the proposed method. Third row shows the ground truth for salient object motion detection.

#### 4.3. Subjective evaluation of the proposed technique

The evaluation mechanism used in this section is based on subjective rating because the prioritization quality assessment is inherently subjective. A group of ten human users were asked to select high-priority images. The participants' ages ranged from 30 to 50 years and they had an average experience of ten years in video surveillance. High-priority images were selected depending on the interestingness from the surveillance point of view. The prioritized images that were manually selected by the participants were used as ground truths for comparison with the prioritized images generated by the proposed and Irgan methods.

The number of matched and unmatched prioritized images was counted between the ground truth and automated method to measure the terms True positive (TP), True Negative (TN) and False Positive (FP). A prioritized image is TP, if the corresponding image is selected by both the experts and the underlying technique. On the other hand, TN is an image that was chosen as high-priority by the technique but not by the experts. FN is an image that was chosen as high-priority by experts but not by the technique. However, it is important to mention that matching does not require the two images to be exactly alike. They were assumed as being the same if they conveyed the same semantic information. Various metrics can be used to evaluate the performance of visual surveillance systems. The most common metrics include *Recall*, *Precision*, and *F-measure* [52-54]. The system of excellent performance is characterized by high values of *Recall*, *Precision*, and *F-measure*. These are the standard metrics used in the pattern recognition community and defined as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (14)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{F - measure} = \frac{1}{\omega \times \frac{1}{\text{Recall}} + (1 - \omega) \times \frac{1}{\text{Precision}}} \quad (16)$$



The *F-measure* is the weighted harmonic mean of the *Recall* and *Precision* [55] as described in equation 16. The parameter  $\omega$  controls the relative importance of *Recall* and *Precision*. In a majority of the surveillance applications *Recall* and *Precision* are given equal importance by setting  $\omega = 0.5$  [56, 57]. Putting  $\omega = 0.5$  in above equation we get

$$F - \text{measure} = 2 \cdot \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (17)$$

Here, *Recall* is the ability of a particular method to correctly identify a video frame as high-priority. *Precision* is the probability that a relevant frame is selected as high priority. On the other hand, *F-measure* is used to reflect the average of both *Recall* and *Precision*. The *F-measure* metric is incorporated because *Recall* and *Precision* are complementary to each other and thus cannot be used alone. For example, a significantly high value of *Precision* can be achieved by selecting only few images as high-priority. Similarly, selecting too many images as high priority can lead to a high *Recall* value. However, a high value of *F-measure* indicates a high value for both *Recall* and *Precision*.

Tables 2 and 3 show the results of the proposed and Irgan methods on the data set mentioned in Table 1. Table 3 shows the *Recall*, *Precision*, and *F-measure* values on single camera videos. In this case, single camera videos were considered from each video sequence. The proposed technique outperforms the other technique by consistently achieving high scores for all the three measures. For these single camera videos, our scheme achieves a high value of *Recall* and *Precision*. Thus, the *F-measure* of our method exceeds that of the Irgan method, thereby illustrating that the proposed method is more effective in case of single-camera videos. However, in dealing with multi-camera videos, Table 4 shows a slightly different behavior. Here, the proposed method achieves a high value of *Precision* because it efficiently selects the semantically relevant camera stream. However, its *Recall* value is significantly low because it picks only the video frames having high salient motion while discarding the non-significant video streams. However, the Irgan method fails to distinguish between the salient and non-salient motion with accuracy. Thus, it chooses and transmits frames that are often of high-priority and occasionally are of low-priority (inclusive error) from all camera streams, which creates redundancy. However, the proposed method efficiently removes redundancy while providing almost the same *F-measure* as the other method. This is feasible and suited to the requirement of WVSNS, where streaming large data is impractical because of energy and bandwidth constraints.



Hence, it can be deduced that the proposed method is suitable for multi-camera surveillance applications in resource-constrained setups.

			[8]			Proposed		
Sr.	Multi-camera video sequence name	No. of cameras	Recall	Precision	F-measure	Recall	Precision	F-measure
1	Sequence 1	1	0.72	0.77	<b>0.74</b>	0.74	0.73	0.73
2	Sequence 2	1	0.71	0.74	0.72	0.79	0.76	<b>0.77</b>
3	Sequence 10	1	0.75	0.79	0.77	0.81	0.82	<b>0.81</b>
4	Sequence 6	1	0.75	0.69	0.72	0.85	0.8	<b>0.82</b>
5	ETI-VS2-AP	1	0.74	0.71	0.72	0.78	0.74	<b>0.76</b>
6	ETI-VS2-BE-19	1	0.80	0.75	0.77	0.83	0.81	<b>0.82</b>
<b>Average</b>			<b>0.75</b>	<b>0.74</b>	<b>0.74</b>	<b>0.80</b>	<b>0.78</b>	<b>0.79</b>

Table 3: Recall, Precision, and F-measure achieved using the mentioned techniques on single camera video sequences.

			[8]			Proposed		
Sr.	Multi-camera video sequence name	No. of cameras	Recall	Precision	F-measure	Recall	Precision	F-measure
1	Sequence 1	10	0.87	0.75	<b>0.81</b>	0.74	0.8	0.77
2	Sequence 2	10	0.84	0.74	<b>0.79</b>	0.71	0.86	0.78
3	Sequence 10	10	0.89	0.79	<b>0.84</b>	0.77	0.9	0.83
4	Sequence 6	10	0.88	0.69	<b>0.77</b>	0.7	0.79	0.74
5	ETI-VS2-AP	4	0.82	0.69	<b>0.75</b>	0.66	0.82	0.73
6	ETI-VS2-BE-19	4	0.78	0.68	<b>0.72</b>	0.64	0.8	0.71
<b>Average</b>			<b>0.85</b>	<b>0.72</b>	<b>0.77</b>	<b>0.70</b>	<b>0.83</b>	<b>0.76</b>

Table 4: Recall, Precision, and F-measure achieved using the mentioned techniques on multi-camera video sequences.

#### 4.4. Effectiveness of the proposed method

This section focuses on the effectiveness of the proposed scheme with respect to the network lifetime maximization problem. This is achieved through the comparative analysis of bandwidth and energy utilization.

##### a) Bandwidth utilization

This section highlights the transmission energy saving aspect of the proposed framework. Prioritized Packet Rate (PPR) is used to measure the efficiency of the proposed method in terms of bandwidth utilization and transmission energy saving. PPR is the ratio of prioritized information; it reflects the bandwidth utilization of the transmission path. Assume that  $P_{i,j}^{t,h}$  is  $i^{\text{th}}$  high-priority image block, captured at time  $t$  by camera sensor  $j$ . Here,  $h$  represents the high-priority macro-blocks. Then, the PPR is defined as

$$\text{PPR} = \frac{\sum_{k=1}^{N^{\text{PH}}} \sum_{i=1}^n P_{i,j}^{t,h}}{\sum_{k=1}^N \sum_{i=1}^n P_{i,k}^t} \quad (18)$$

where  $n$  represents the number of macro-blocks in an image and  $N$  is the total number of camera sensors.  $N^{\text{PH}}$  is the total number of high-priority cameras. Equation 18 calculates the ratio of the sum of high-priority macro-blocks to the total number of macro-blocks obtained from the multi-camera generated frames. To understand the PPR-based comparative analysis, visual results of priority schemes are provided in figure 10. In this figure, the priority measure of each image block is estimated using the proposed and Irgan prioritization schemes. The Irgan prioritization scheme includes five measures: edge measure, entropy measure, reproducibility measure, delta-entropy measure, and delta-edge measure. For evaluation of the image prioritization schemes *Recall*, *Precision*, and *F-measure* are used (as explained in the above section). These measures are based on True positive (TP), True Negative (TN) and False Positive (FP) where TN refers to the pixels selected by the technique as high priority but not by the experts.

As can be seen in figure 10, the image blocks corresponding to objects such as fountain, car, and pole have the highest edge measures. Although the edge measure detects possible objects in the scene, it suffers from an inclusive error by considering various non-important objects as important. In the case of entropy measure, the macro-blocks corresponding to the background have lower entropy values, whereas foreground objects have relatively higher entropy values. Furthermore, it can be seen that the performance of delta-entropy measure is worse than the entropy measure. However, the delta-edge measure outperforms other Irgan prioritization measures. It can be noticed from Figure 10, that there are high values of TN in the case of Irgan measures as compared to the proposed prioritization measure. Thus, a significant portion of image that is not important from a surveillance point of view is considered as high priority. This increases the transmission burden on the wireless network as illustrated in Figure 11, which

shows the PPR results on a segment (frame 1222 to frame 1296) of the *fountain02.mpeg* video in which a car moves across the road as described in Table 2. It can be deduced from Figure 11 that the proposed prioritization scheme achieves the best performance by transmitting a maximum amount of the salient object pixels with the minimum bandwidth.

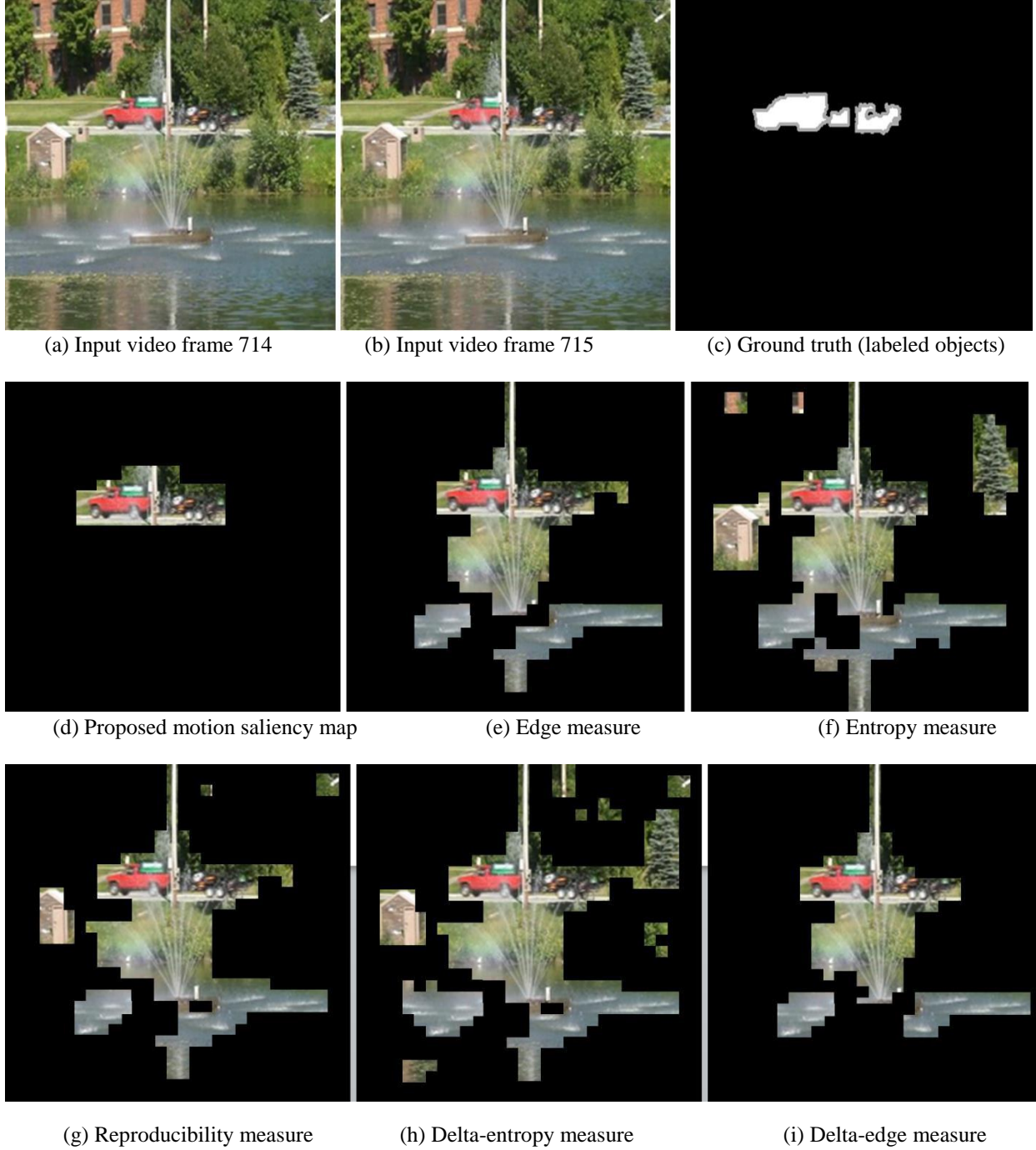


Fig. 10: Visual illustration of packet priority measure schemes in the sample frame sequence of the *fountain02.mpeg* video.

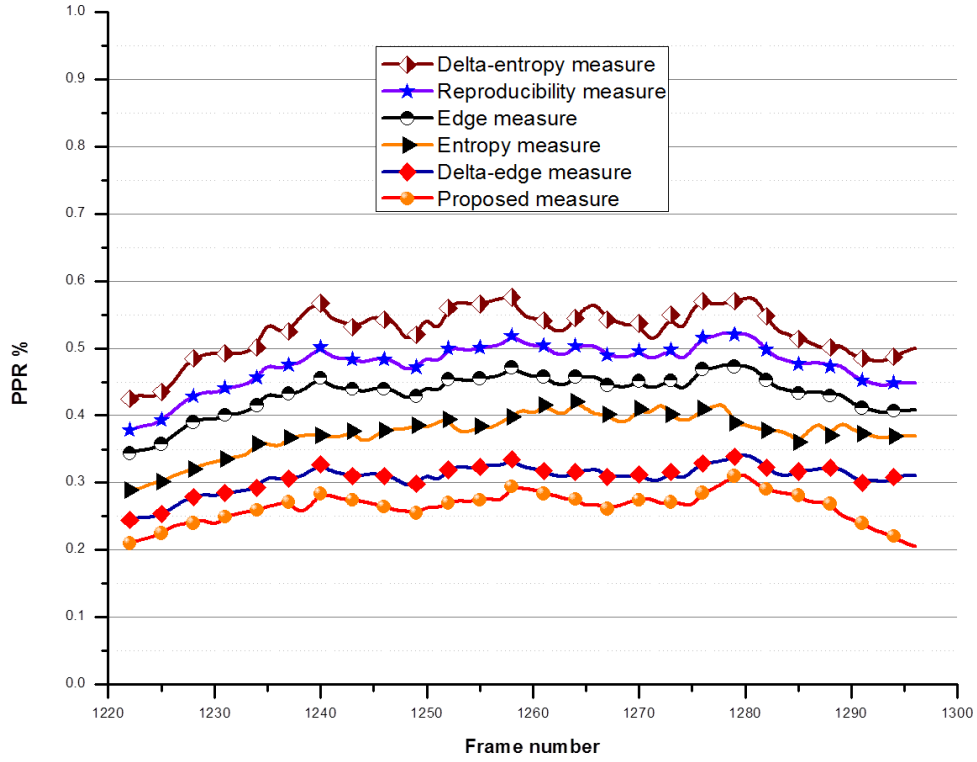


Fig. 11: PPR values for each method on the “Sequence 10” video shot by ten cameras.

#### b) Computational energy and time analysis

In this section, computational energy and time are analyzed for a number of well-known prioritization measures. In order to render the article self-contained, a brief description of these measures is given as follows:

- i. Entropy measure: the probabilistic information measure of a macro-block that can be calculated by its normalized gray-scale histogram image.
- ii. Edge measure: a metric commonly used in prioritization; it is based on the assumption that if there is an object it should have edges. Here, it is calculated using a Haar filter.
- iii. Reproducibility measure: it assigns weights to each macro-block by finding some of the difference of its mean value with its eight neighboring blocks. This reproducibility-measure assigns high weights to image

blocks containing objects, which makes it an appropriate candidate for measuring the priority value of an image.

- iv. Hybrid measure: it includes delta edge and delta entropy measure, obtained by combining reproducibility with edge and entropy respectively. For further details, readers can refer [8].
- v. Robust salient motion detection measure (RSMD) [34]: it prioritizes visual images by detecting salient motion in complex environments. Salient motion is detected by combining temporal filtered motion field and temporal difference.
- vi. Abrupt motion tracking measure (AMT) [35]: in this measure surveillance scenes are prioritized using visual saliency based particle filter tracker. It computes saliency model using a top-down scheme which is target-driven.
- vii. Salient region detection measure (SRD) [36]: it utilizes temporal variations in the flow field of a crowd scene to detect salient regions. Optical flow is employed to estimate the velocity field at each point in the scene. Based on velocity field temporal variations, salient motion regions are highlighted.

Table 5 shows the average execution time and energy consumption of an  $8 \times 8$  macro-block measured for a number of image prioritization methods. The computational cost of each measure is provided in terms of CPU cycles, program memory, data memory, time, and energy. In terms of time and energy consumption, the results are in favor of the proposed priority measure because it detects the motion of salient objects using integral-image based temporal gradients. In integral-based temporal gradients, the calculation of the integral-image is the main computational step; following this calculation, only four array-references are required for the computation of a specific block. This accelerates the analysis of an integral-image block and makes it independent of the macro-block size. Therefore, it can be concluded that the proposed priority measure is more efficient than other state-of-the-art measures for resource constrained WVSNS.

Methods	CPU cycles	Program memory (bytes)	Data memory (bytes)	Network buffer (bytes)	Time (ms)	Energy ( $\mu$ J)
Entropy	8270	206	454	64	1.03	3.1
Edge	2800	126	72	64	0.35	1.05
Reproducibility	1820	514	154	4480	0.23	0.68
Delta-entropy	10090	720	608	4544	1.26	3.78
Delta-edge	4620	640	226	4544	0.58	1.73
RSMD [34]	18500	1256	1280	6880	1.84	4.91

<b>AMT [35]</b>	25110	1656	1536	7480	2.70	5.85
<b>SRD [36]</b>	11250	930	850	6440	1.30	4.70
<b>Proposed method</b>	1280	570	125	4480	0.20	0.31

Table 5: Average computational cost (time and energy) of the prioritization measures for a macro-block.

## 5. Conclusion

In surveillance applications, visual data transmission over wireless sensors networks consumes more energy resources than the transmission of scalar data. In order to maximize network lifetime, surveillance applications require efficient extraction of succinct information by eliminating the redundant and non-informative data. In this paper, we have proposed a cost-effective image prioritization framework for wireless sensor networks. Novel integral-image based temporal gradients are introduced to calculate the salient motion in videos efficiently. This enables visual frames to be categorized as high-priority or low-priority. The camera sensors send their priority values to the SN node before sending data to it. If the portion of the event area covered by a particular camera sensor is significant and the value of the calculated priority is greater than existing priority camera sensors, then that camera is assigned a reliable channel otherwise the data is dropped. This guarantees the lossless and timely transmission of salient / high-priority data.

The performance of monitoring applications depends on salient activity coverage that requires several camera sensors with overlapping FoVs. However, numerous cameras with overlapping FoVs results is unnecessary and redundant visual data transmission to the SN. The proposed framework prioritizes multi-camera data to minimize the amount of data that is transmitted significantly. This results in the loss of some monitoring quality; however, it is expected that surveillance applications can tolerate some loss of monitoring quality due to the elimination of low-priority data. Moreover, the proposed framework has the potential to prolong the lifetime of the WWSN while still providing a significant monitoring quality comparable to the case where all the cameras in the vicinity are transmitting all their data.

The architecture operates entirely on integer-based arithmetic and requires considerably lesser memory than traditional monitoring applications, making it a feasible for a limited resource wireless surveillance networks. Extensive numerical simulations show that the proposed framework can support a considerably longer network lifetime when compared to the existing prioritization schemes optimized for the WWSN assisted monitoring

applications. It significantly reduces transmission costs. In addition, the performance of the proposed integral-image based temporal gradients is considerably better than the exiting measures such as delta-edge, and delta-entropy.

## Acknowledgements

This work was supported by the Industrial Strategic Technology Development Program, 10041772, The Development of an Adaptive Mixed-Reality Space Based on Interactive Architecture, funded by the Ministry of Science, ICT & Future Planning (MSIP).

## References:

- [1] I.F. Akyildiz, T. Melodia, K.R. Chowdury, Wireless multimedia sensor networks: A survey, *Wireless Communications, IEEE*, 14 (2007) 32-39.
- [2] M. Casares, M.C. Vuran, S. Velipasalar, Design of a wireless vision sensor for object tracking in wireless vision sensor networks, in: *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, IEEE, 2008, pp. 1-9.
- [3] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, A survey on sensor networks, *Communications magazine, IEEE*, 40 (2002) 102-114.
- [4] L. Šerić, D. Stipaničev, M. Štula, Observer network and forest fire detection, *Information Fusion*, 12 (2011) 160-175.
- [5] Y.-W. Hong, W.-J. Huang, F.-H. Chiu, C.-C. Kuo, Cooperative communications in resource-constrained wireless networks, *Signal Processing Magazine, IEEE*, 24 (2007) 47-57.
- [6] B. Sahoo, S. Rath, D. Puthal, Energy Efficient Protocols for Wireless Sensor Networks: A Survey and Approach, *International Journal of Computer Applications (0975–8887)*, 44 (2012).
- [7] H.E. Pashler, *The psychology of attention*, MIT press, 1999.
- [8] K. Irgan, C. Ünsalan, S. Baydere, Low-cost prioritization of image blocks in wireless sensor networks for border surveillance, *Journal of Network and Computer Applications*, (2013).
- [9] Z. Sun, P. Wang, M.C. Vuran, M.A. Al-Rodhaan, A.M. Al-Dhelaan, I.F. Akyildiz, BorderSense: Border patrol through advanced wireless sensor networks, *Ad Hoc Networks*, 9 (2011) 468-477.
- [10] L. Zhou, A.V. Vasilakos, N. Xiong, Y. Zhang, S. Lian, Scheduling security-critical multimedia applications in heterogeneous networks, *Computer Communications*, 34 (2011) 429-435.
- [11] Y. Ye, S. Ci, A. Katsaggelos, Y. Liu, Y. Qian, *Wireless Video Surveillance: A Survey*.
- [12] Y. Nam, S. Rho, J.H. Park, Intelligent Video Surveillance System: 3-tier context-aware surveillance system with metadata, *Multimedia Tools and Applications*, 57 (2012) 315-334.
- [13] R. Tan, G. Xing, B. Liu, J. Wang, X. Jia, Exploiting data fusion to improve the coverage of wireless sensor networks, *Networking, IEEE/ACM Transactions on*, 20 (2012) 450-462.
- [14] O. Demigha, W. Hidouci, T. Ahmed, On energy efficiency in collaborative target tracking in wireless sensor network: a review, (2012).
- [15] T.-H. Tsai, C.-Y. Lin, Exploring contextual redundancy in improving object-based video coding for video sensor networks surveillance, *Multimedia, IEEE Transactions on*, 14 (2012) 669-682.
- [16] B. Dieber, C. Micheloni, B. Rinner, Resource-aware coverage and task assignment in visual sensor networks, *Circuits and Systems for Video Technology, IEEE Transactions on*, 21 (2011) 1424-1437.
- [17] M. Senouci, A. Mellouk, L. Oukhellou, A. Aissani, An Evidence-Based Sensor Coverage Model, (2012).

- [18] Y. Wang, G. Cao, On full-view coverage in camera sensor networks, in: INFOCOM, 2011 Proceedings IEEE, IEEE, 2011, pp. 1781-1789.
- [19] Y. Wang, G. Cao, Barrier coverage in camera sensor networks, in: Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing, ACM, 2011, pp. 12.
- [20] M.P. Johnson, A. Bar-Noy, Pan and scan: Configuring cameras for coverage, in: INFOCOM, 2011 Proceedings IEEE, IEEE, 2011, pp. 1071-1079.
- [21] A. Newell, K. Akkaya, Distributed collaborative camera actuation for redundant data elimination in wireless multimedia sensor networks, *Ad Hoc Networks*, 9 (2011) 514-527.
- [22] N. Tezcan, W. Wang, Self-orienting wireless multimedia sensor networks for occlusion-free viewpoints, *Computer networks*, 52 (2008) 2558-2567.
- [23] J. Chen, J. Li, S. He, Y. Sun, H.-H. Chen, Energy-efficient coverage based on probabilistic sensing model in wireless sensor networks, *Communications Letters, IEEE*, 14 (2010) 833-835.
- [24] S. Kadhe, S. Thaskani, M.G. Chandra, B. Adiga, Reliable data transmission in sensor networks using compressive sensing and real expander codes, in: Communications (NCC), 2012 National Conference on, IEEE, 2012, pp. 1-5.
- [25] T. Ma, M. Hempel, D. Peng, H. Sharif, A Survey of Energy-Efficient Compression and Communication Techniques for Multimedia in Resource Constrained Systems, (2012).
- [26] F. Li, Y. Gao, ROI-based error resilient coding of H. 264 for conversational video communication, in: Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International, IEEE, 2011, pp. 1719-1723.
- [27] J. Lei, M. Wu, K. Feng, C. Hu, C. Hou, Multilevel region of interest guided bit allocation for multiview video coding, *Optik-International Journal for Light and Electron Optics*, 125 (2014) 39-43.
- [28] K. Lam, Y. Kaneoke, A. Gunji, H. Yamasaki, E. Matsumoto, T. Naito, R. Kakigi, Magnetic response of human extrastriate cortex in the detection of coherent and incoherent motion, *Neuroscience*, 97 (2000) 1-10.
- [29] W. Hu, T. Tan, L. Wang, S. Maybank, A survey on visual surveillance of object motion and behaviors, *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34 (2004) 334-352.
- [30] S. Zahurul, R. Jidin, N.M. Din, S. Zahidul, An Improved and Secure Motion Detection Surveillance System in UNIX, *International Journal of Applied Electronics in Physics & Robotics*, 1 (2013) 13-17.
- [31] J. Dou, J. Li, Modeling the background and detecting moving objects based on Sift flow, *Optik-International Journal for Light and Electron Optics*, 125 (2014) 435-440.
- [32] S.-C. Huang, F.-C. Cheng, Motion detection with pyramid structure of background model for intelligent surveillance systems, *Engineering Applications of Artificial Intelligence*, 25 (2012) 1338-1348.
- [33] K.-T. Hu, J.-J. Leou, H.-H. Hsiao, Spatiotemporal saliency detection and salient region determination for H. 264 videos, *Journal of Visual Communication and Image Representation*, 24 (2013) 760-772.
- [34] Y.-L. Tian, A. Hampapur, Robust salient motion detection with complex background for real-time video surveillance, in: Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on, IEEE, 2005, pp. 30-35.
- [35] Y. Su, Q. Zhao, L. Zhao, D. Gu, Abrupt motion tracking using a visual saliency embedded particle filter, *Pattern Recognition*, (2013).
- [36] M. Lim, C. Chan, D. Monekosso, P. Remagnino, Detection of salient regions in crowded scenes, *Electronics Letters*, 50 (2014) 363-365.
- [37] A. Marcus, O. Marques, An eye on visual sensor networks, *Potentials, IEEE*, 31 (2012) 38-43.
- [38] F.C. Crow, Summed-area tables for texture mapping, in: ACM SIGGRAPH Computer Graphics, ACM, 1984, pp. 207-212.



- [39] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, IEEE, 2001, pp. I-511-I-518 vol. 511.
- [40] Y.-T. Chen, C.-S. Chen, Fast human detection using a novel boosted cascading structure with meta stages, Image Processing, IEEE Transactions on, 17 (2008) 1452-1464.
- [41] H. Lee, H. Byun, Temporal gradient pattern for the near-duplicate video clustering, in: Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, IEEE, 2010, pp. 3186-3189.
- [42] B.W. Tatler, The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions, Journal of Vision, 7 (2007).
- [43] L. Duan, C. Wu, J. Miao, A.C. Bovik, Visual Conspicuity Index: Spatial Dissimilarity, Distance, and Central Bias, Signal Processing Letters, IEEE, 18 (2011) 690-693.
- [44] A.O. Bicen, V.C. Gungor, O.B. Akan, Delay-sensitive and multimedia communication in cognitive radio sensor networks, Ad Hoc Networks, 10 (2012) 816-830.
- [45] I.F. Akyildiz, T. Melodia, K.R. Chowdhury, A survey on wireless multimedia sensor networks, Computer networks, 51 (2007) 921-960.
- [46] R. Umar, A.U. Sheikh, A comparative study of spectrum awareness techniques for cognitive radio oriented wireless networks, Physical Communication, (2012).
- [47] O. Hussain, F. Hussain, S. Parvin, S. Han, B. Tian, E. Chang, Cognitive radio network security: A survey, (2012).
- [48] I.F. Akyildiz, B.F. Lo, R. Balakrishnan, Cooperative spectrum sensing in cognitive radio networks: A survey, Physical Communication, 4 (2011) 40-62.
- [49] W. Zhang, R. Mallik, K. Letaief, Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks, Wireless Communications, IEEE Transactions on, 8 (2009) 5761-5766.
- [50] R.A. Abrams, S.E. Christ, Motion onset captures attention, Psychological Science, 14 (2003) 427-432.
- [51] N. Goyette, P.-M. Jodoin, F. Porikli, J. Konrad, P. Ishwar, Changedetection. net: A new change detection benchmark dataset, in: Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, IEEE, 2012, pp. 1-8.
- [52] N. Lazarevic-McManus, J. Renno, G. Jones, Performance evaluation in visual surveillance using the F-measure, in: Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, ACM, 2006, pp. 45-52.
- [53] S.K. Pal, A. Petrosino, L. Maddalena, Handbook on Soft Computing for Video Surveillance, CRC Press, 2012.
- [54] N. Lazarevic-McManus, J. Renno, D. Makris, G.A. Jones, An object-based comparative methodology for motion detection based on the  $F$ -Measure, Computer Vision and Image Understanding, 111 (2008) 74-85.
- [55] J. Makhoul, F. Kubala, R. Schwartz, R. Weischedel, Performance measures for information extraction, in: Proceedings of DARPA Broadcast News Workshop, 1999, pp. 249-252.
- [56] A. Karimaa, Efficient video surveillance: performance evaluation in distributed video surveillance systems, Video Surveillance. InTech, (2011).
- [57] J.W. Davis, A.M. Morison, D.D. Woods, An adaptive focus-of-attention model for video surveillance and monitoring, Machine Vision and Applications, 18 (2007) 41-64.